Leading from Incidents

Nora Jones LeadDev NY 2023



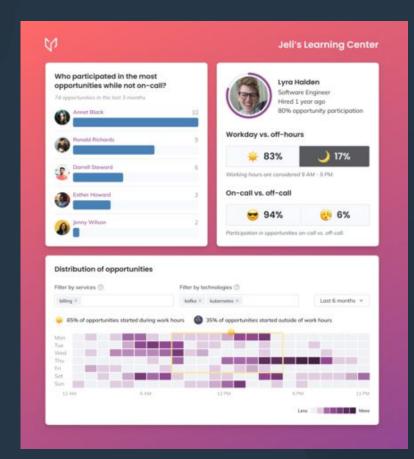
Don't be better alone. Bring others with you.





Venture

GamesBeat Data Pipeline Low-Code/No-Code Summit





We're all here because:

- The industry is changing. We've been through a lot in the last 3 years.
- There's more to learn from the incidents we have.
- You're here to share and learn and collaborate about the better ways.
- Everyone is trying to do more with less right now.



Learning from incidents...

Learning from incidents asks not why it happened, but:
 "Why did it make sense for us to do it this way?"



Learning from incidents...

So I'm going to ask that of the industry...

"Why did it make sense for us to do it this way?"

So that we can be better, in 2023 and beyond.



Part 1. How we got here

Part 2. A tale of two incident investigations

Part 3. What's next



Part 1 - How we got here



Let's talk about why learning from incidents is important.

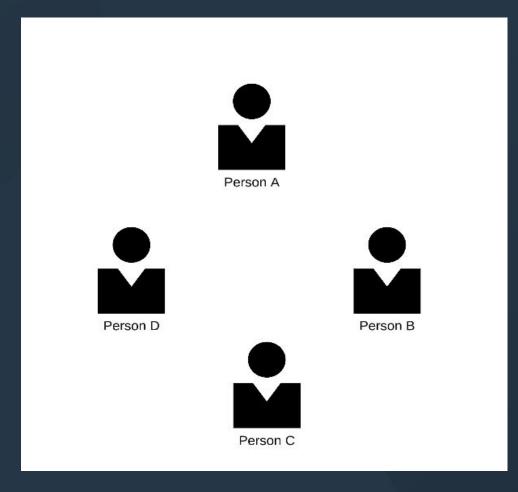


The genesis of expertise amongst individuals



In the beginning of a company...









Person A

"I'd like to use \$NEW_TECHNOLOGY to make this project go faster. It will set us up for success in the long run"



Person D



Person B







Person A





"Ok. If it gets us there faster and more reliably let's do it."





Person A



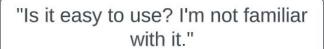
Person D



Person B



Person C







"I'll teach you all. It's fairly straightforward"















Person B











Person B







"I'll teach you all. It's fairly straightforward"

Person A



Person D



Person B









"I'll teach you all. It's fairly straightforward"

Person A











Months pass...

The feature that \$NEW_TECHNOLOGY was used for was a big hit (yay!) and generated a lot of demand for the product. Which means...



\$NEW_TECHNOLOGY





Person D





























\$NEW_TECHNOLOGY



















Person C















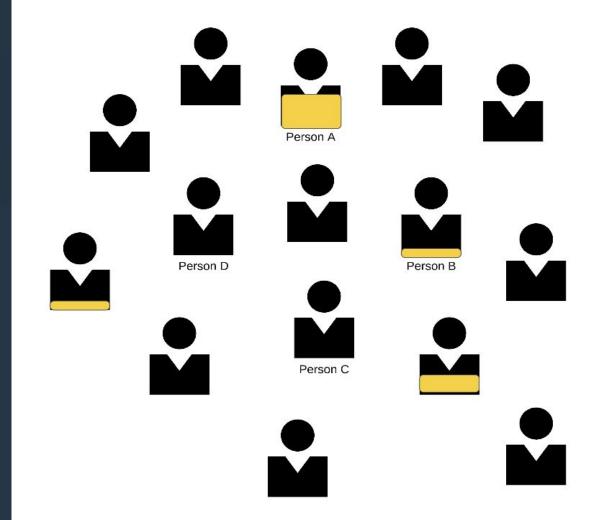


Woods' Theorem:

"As the complexity of a system increases, the accuracy of any single agent's own model of that system decreases rapidly."

-SNAFU Catchers Cycle 1 report (stella.report)

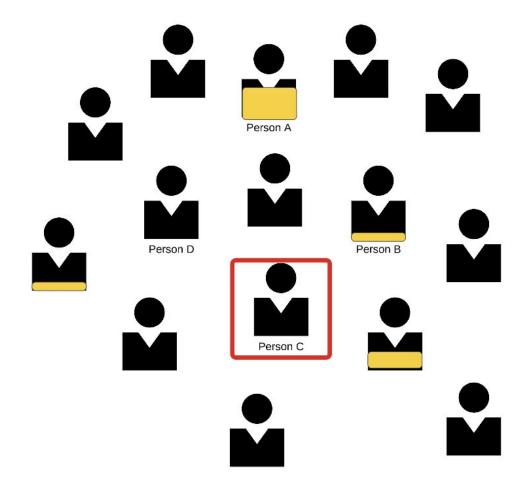




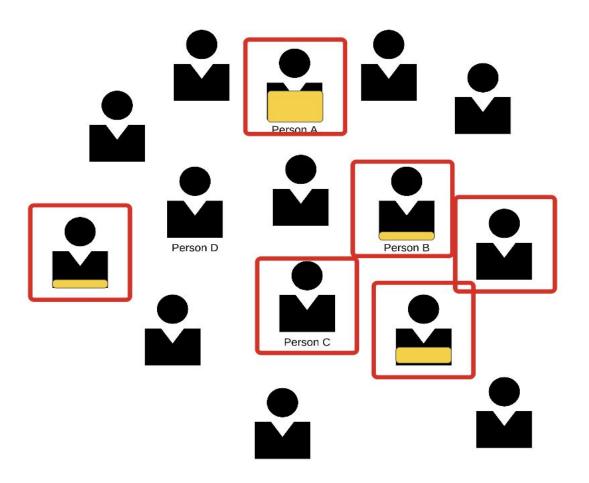


INCIDENT OCCURS









Incident Review happens, but only includes a small group of people and is done hastily. And some experts cannot make the incident review

Management gets high level report with high-level metrics and indicators on the incident.



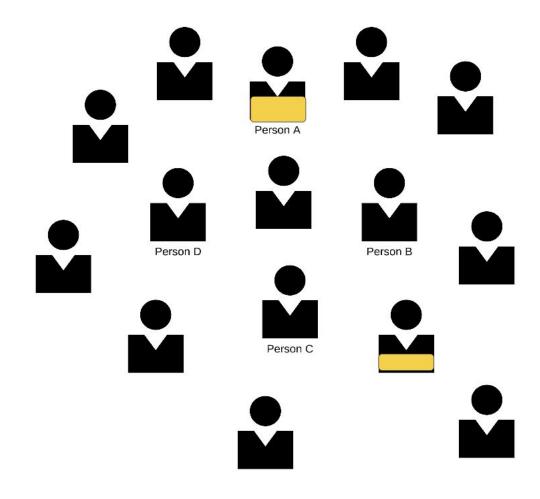
Parties engaged in incident response have different goals and priorities.













Over time, Person A is involved in almost every incident that involves \$NEW_TECHNOLOGY.

They're very valuable to the organization, they have a rich skillset,

As they manage this piece of technology, the organization grows and grows, more people are hired around them with limited knowledge.

They also have a lot going on, they don't have time to teach new people all the ins and outs of \$NEW_TECHNOLOGY. So they keep responding to incidents about it and then moving on to do other things.



- Events that happen since adding piece of technology:
- Company adds more customers
- Company has higher demands
- The scope of Person A's role grows
- Company hires a lot
- They have less time to "teach" people about the technology they implemented
- They just "do the work" on it instead



Multiple, diverse perspectives





You can't meaningfully improve expertise of people without looking at the cognitive work.



You can't meaningfully improve expertise of people without looking at the cognitive work.

Which is, itself, work.



There can be So.
Much.
Awkward.





Part 2 - A tale of two incident reviews.



A seemingly innocuous incident, that didn't have much customer impact, and didn't "deserve" or have time for a thorough postmortem.

A seemingly innocuous incident, that didn't have much customer impact, and didn't "deserve" or have time for a thorough postmortem.

Templated approach

A seemingly innocuous incident, that didn't have much customer impact, and didn't "deserve" or have time for a thorough postmortem.

- Templated approach
- Completed by members of the team most involved in the incident

A seemingly innocuous incident, that didn't have much customer impact, and didn't "deserve" or have time for a thorough postmortem.

- Templated approach
- Completed by members of the team most involved in the incident
- Purpose was to file and report

Post Mortem Template

Summary

Impact

Detection

Resolution

Detailed Summary

Contributing Factors

Timeline

What went well

What went wrong

How we got lucky

Action Items



Post Mortem Template

Summary

Impact

Detection

Resolution

Detailed Summary

Contributing Factors

Timeline

What went well

What went wrong

How we got lucky

Action Items

What don't we know:

- How much time was spent on this document
- Who read it
- Who attended the meeting
- What the purpose of it was
- What was difficult or easy about handling the incident?



Investigation #1: Summary

A change was made to some of the search infrastructure tooling that updates which search index to look at each day.

We were able to recreate the issue and soon after it had been reported, engineering was able to track down the problem and fix the bad configuration.

Due to the nature of the bug, no pre-processing work had to be done for search to be functional again.



Investigation #1: Detailed Summary

The search index is split into collections, one large collection for older [SKUs] and 2-3 collections for recent [SKUs], one per day, for newly created [SKUs].

Each collection can be individually submitted for reading or writing. Since search queries typically read from all collections from a given vendor.

We employ a feature called collection read alias which lets us query from many collections using one name. As new daily collections are created (typically done 2 days in advance), they are added to the read alias.

Due to a bug in our tooling introduced a week before, newly created daily collections were not being added to the read alias. Since collections are created and added to the alias 2 days in advance, the first collection to be affected by the bug was 3 days ago.



Incident #1 Timeline

December 4, 2022	Bug is introduced in python tooling
December 4, 2022	The 6 December collection is created and not added to the alias
December 5, 2022	Newly created SKUs start going into the 6 December alias
December 6, 2022, 2:30 PM	Incident channel starts based on customer reports Search team is paged and begins work
2:31 PM	Issue with elevated error rates is red herring
2:31 PM	Borked broker potential cause
2:33 PM	Borked broker ruled out
2:34 PM	Misconfiguration in read alias identified as the cause
2:36 PM	Configuration is fixed
2:36 PM	Fix is confirmed
2:37 PM	Fix is applied to index file



Incident #1: Impact

All [e-commerce] users had a degraded search experience for the 20 minutes over which the incident took place.



Incident #1: Contributing Factors

- Lack of type system or static analysis in code
- Alert that would have detected this is also broken



Incident #1: Contributing Factors

- <u>Lack of</u> type system or static analysis in code
- Alert that would have detected this is also broken



Incident #1: What went well

Fairly quick fix one the search team staged



Incident #1: What went well

• Fairly quick fix one the search team staged



Incident #1: What went wrong

• Alert that should have caught the misconfiguration was also broken



Incident #1: What went wrong

• Alert that <u>should have</u> caught the misconfiguration was also broken



Incident #1: How we got lucky

• Happened during a low traffic time, and not many users were impacted



Incident #1: How we got lucky

• Happened <u>during a low traffic</u> time, and not many users were impacted



Incident #1: Action Items

- Fix the bug preventing the read alias from being updated
- Fix the alert that should have fired
- Check all our alerts to make sure that they fire



Incident #1: Action Items

- Fix the bug preventing the read alias from being updated
- Fix the alert that should have fired
- Check all our alerts to make sure that they fire



What questions do you have about incident #1?



It's always a little more complex than it appears

 Conducted by an engineer outside of the main team, fresh perspective

- Conducted by an engineer outside of the main team, fresh perspective
- Included evidence from previous deploys (some well before the event)

- Conducted by an engineer outside of the main team, fresh perspective
- Included evidence from previous deploys (some well before the event)
- Included teams impacted, but not previously involved in the timeline

- Conducted by an engineer outside of the main team, fresh perspective
- Included evidence from previous deploys (some well before the event)
- Included teams impacted, but not previously involved in the timeline
- Informed new team members about system dynamics

- Conducted by an engineer outside of the main team, fresh perspective
- Included evidence from previous deploys (some well before the event)
- Included teams impacted, but not previously involved in the timeline
- Informed new team members about system dynamics
- Purpose was to engage the audience and to be read

How We Got Here "HOWIE" Guide

Background of the document Key Takeaways

Who responded Triggering Event

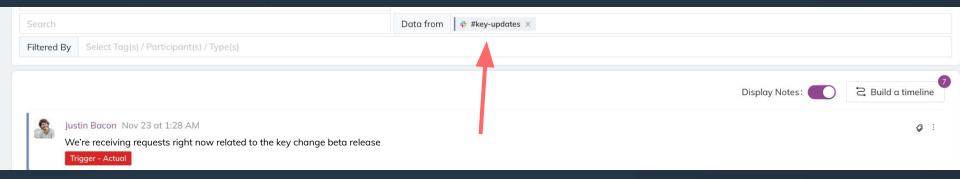
Who investigated Contributors/Enablers

Executive Summary Mitigators

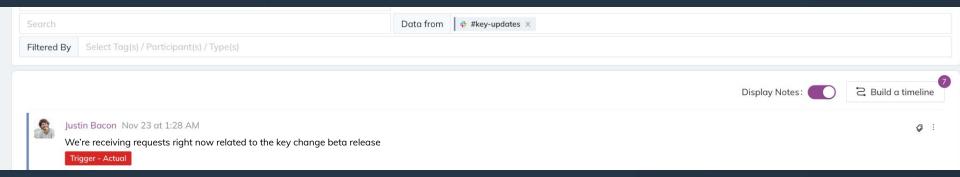
Customer+Employee Impact Difficulties during handling

Follow-up items

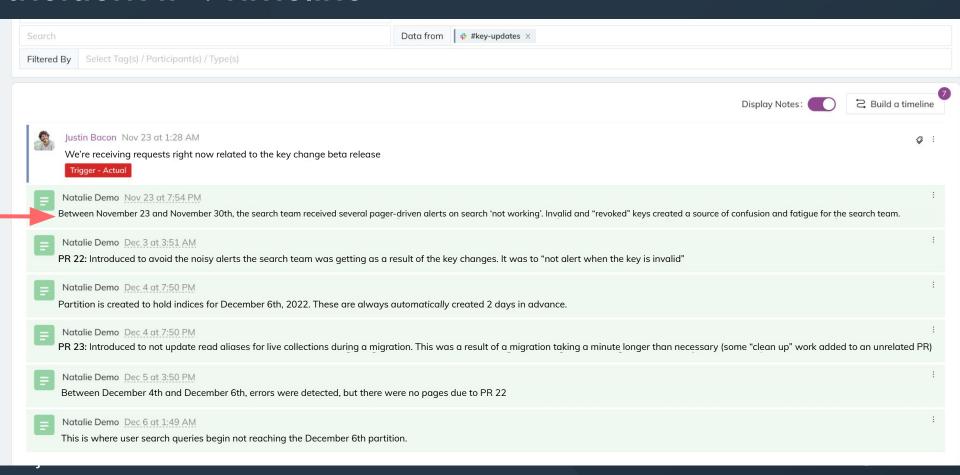


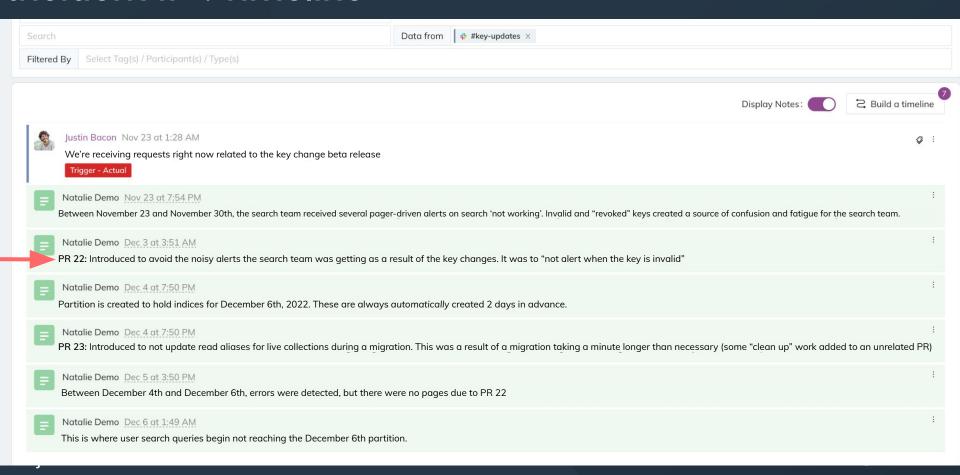


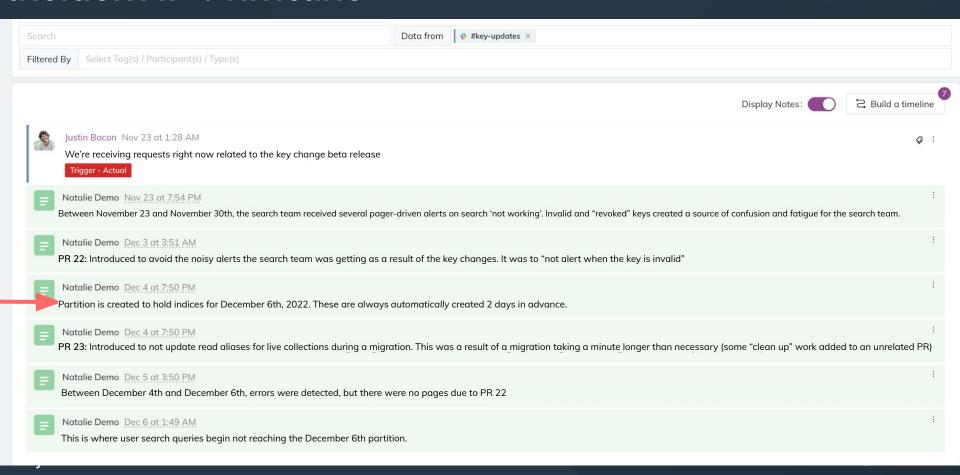


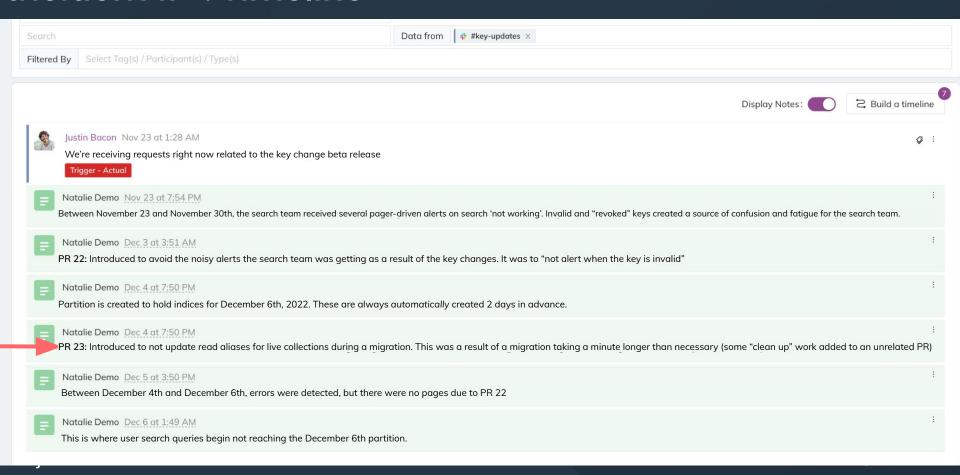


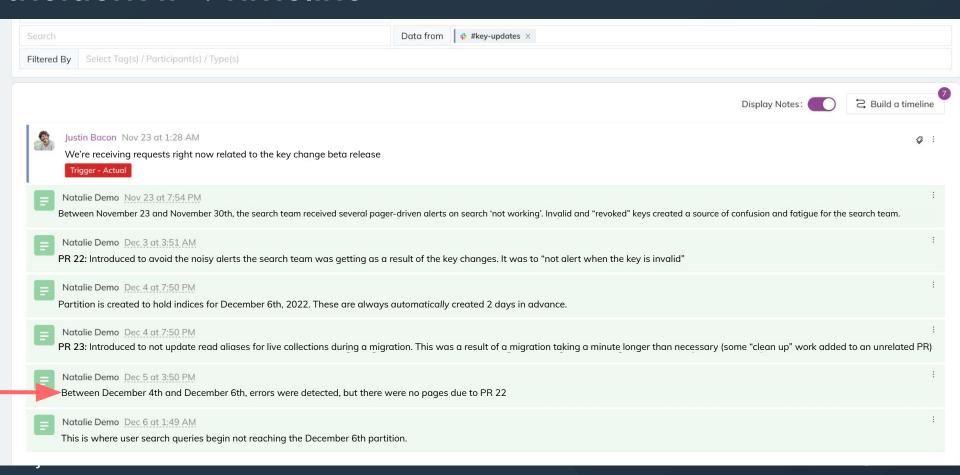


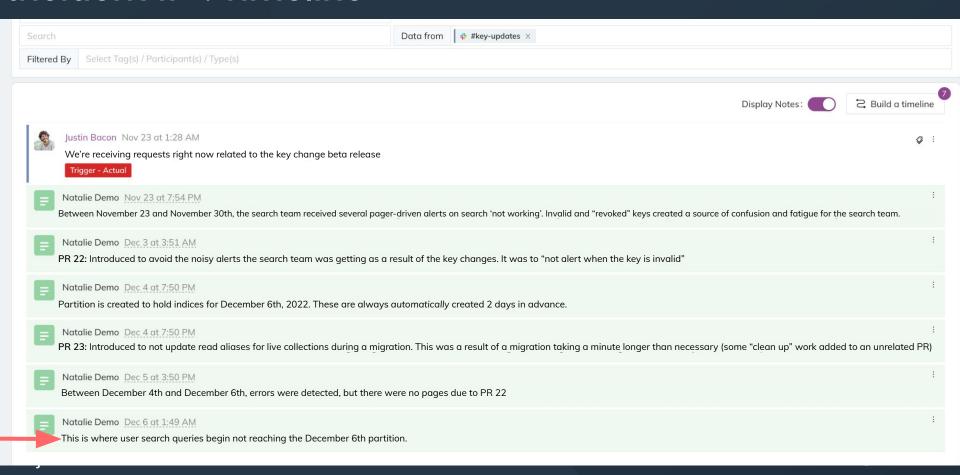


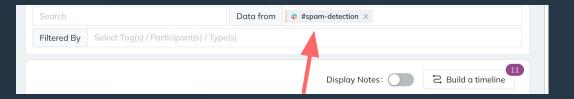




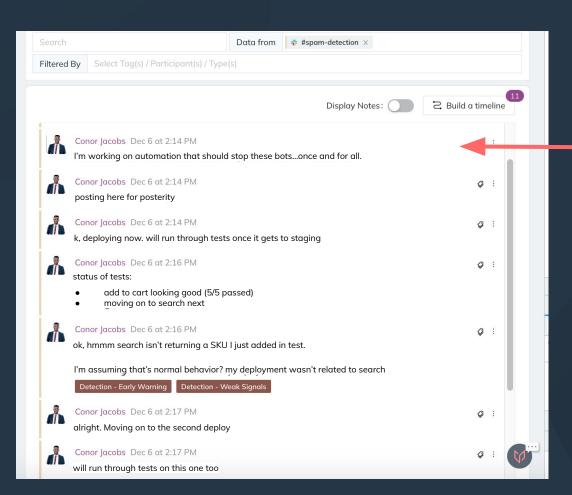




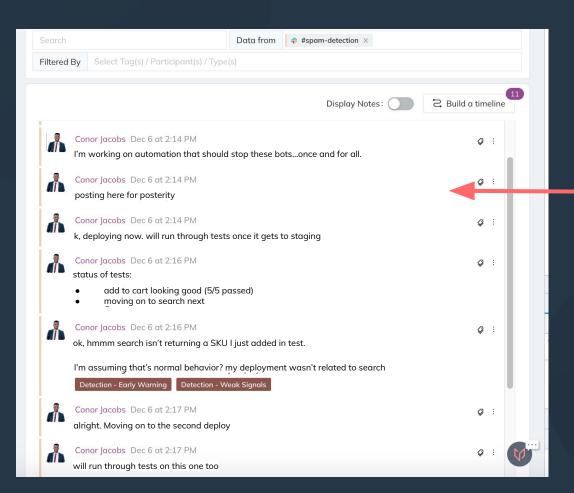




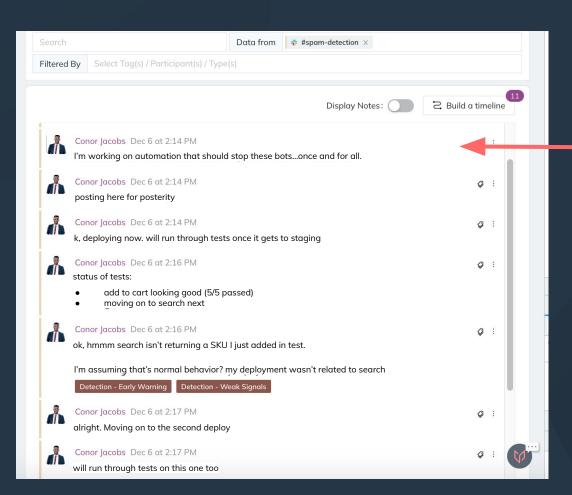




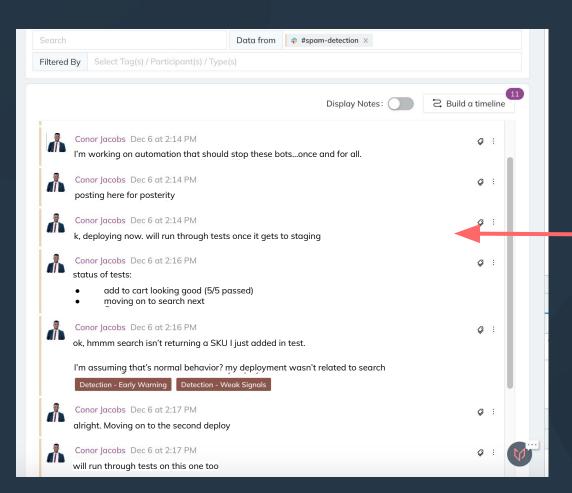




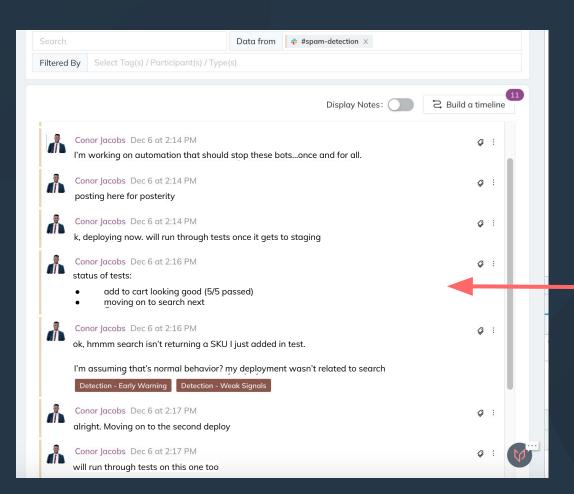




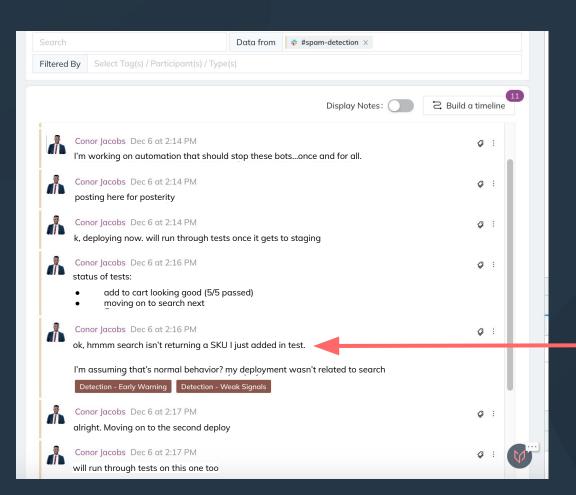




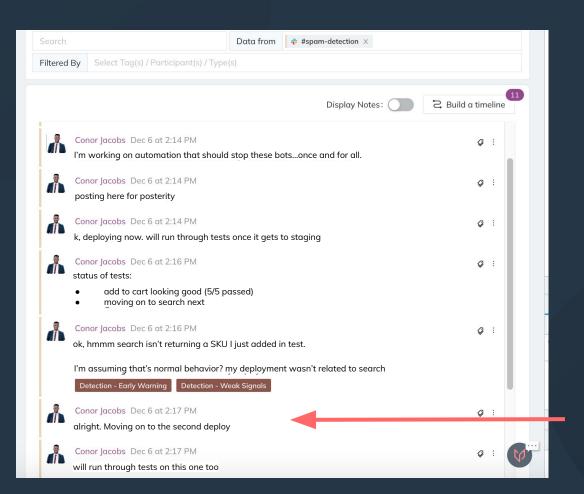




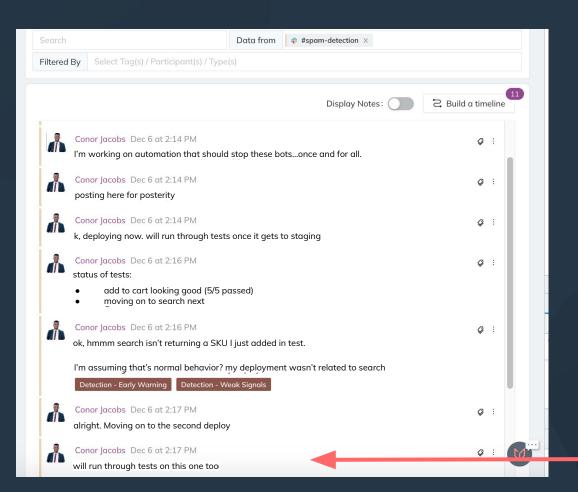




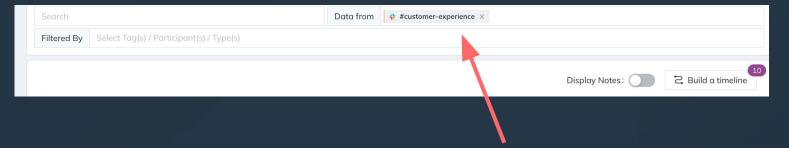




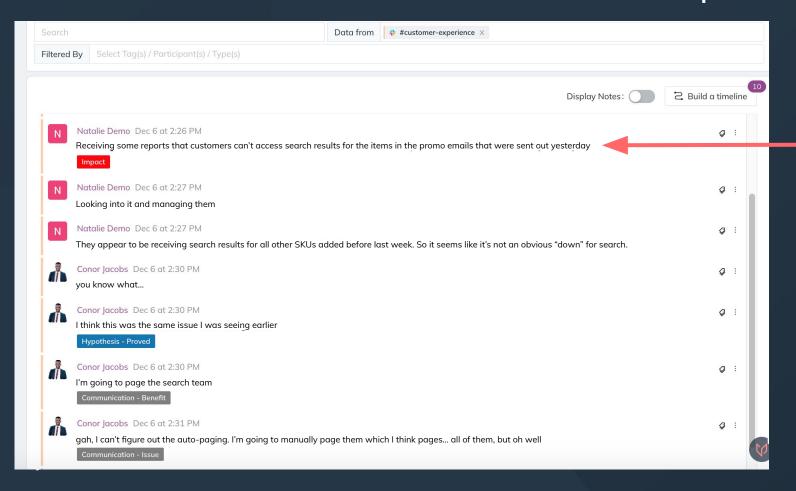




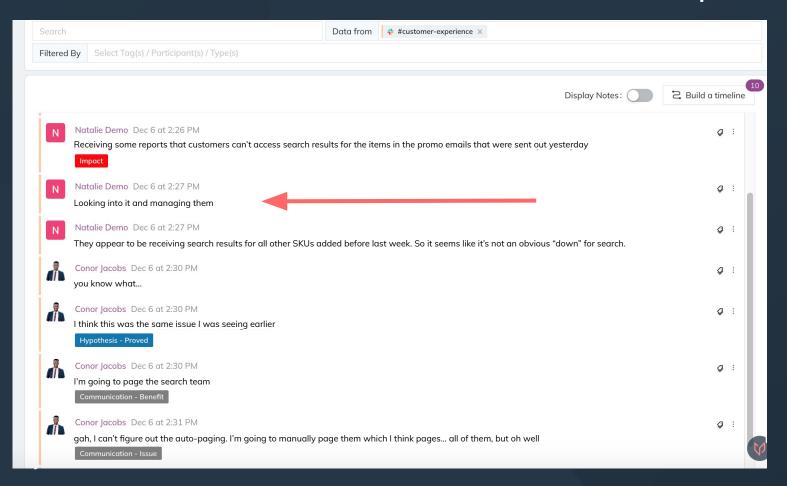




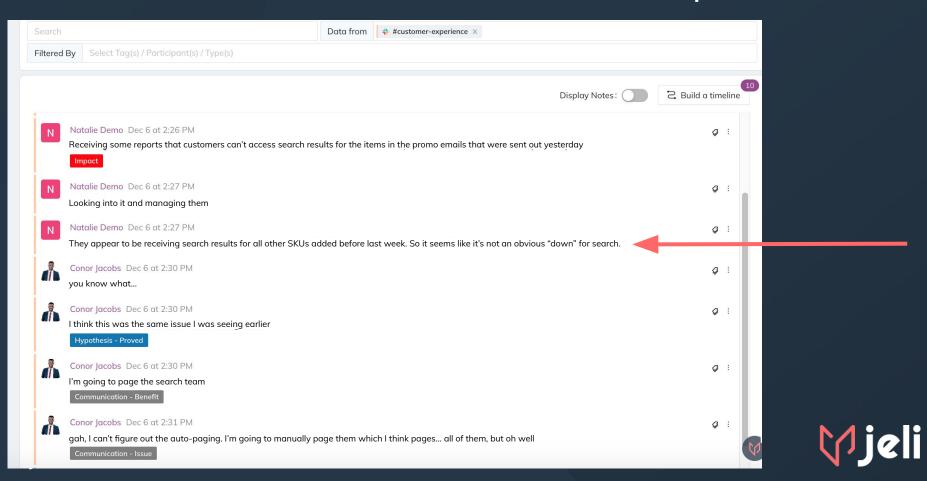


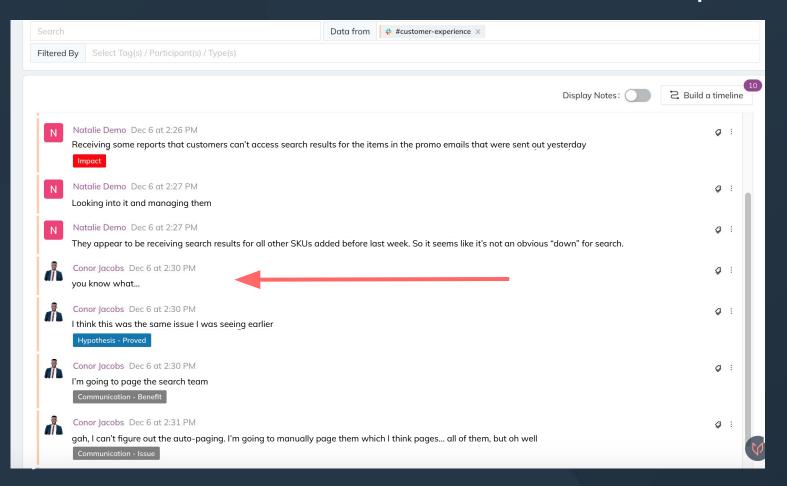




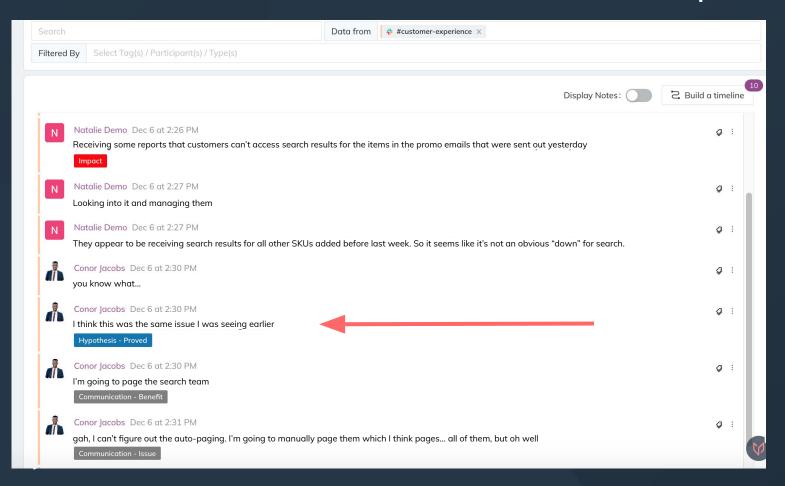




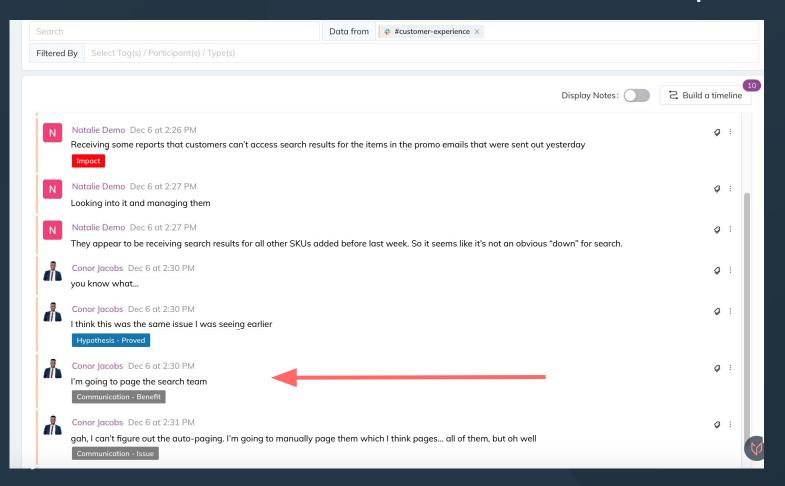




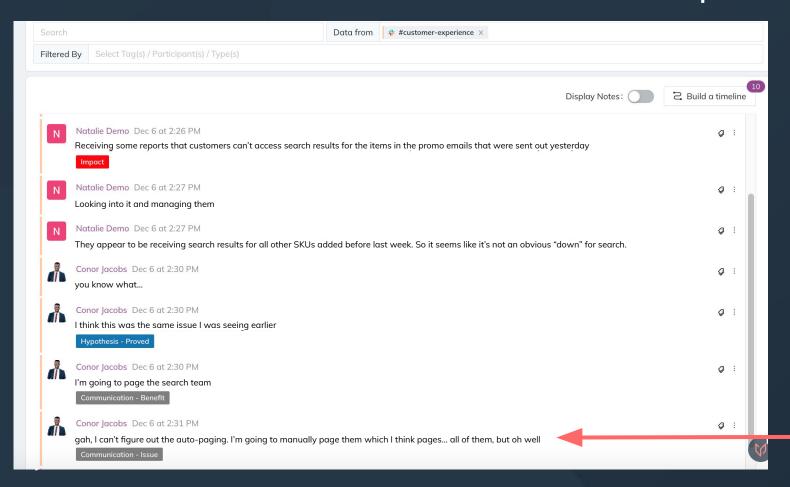




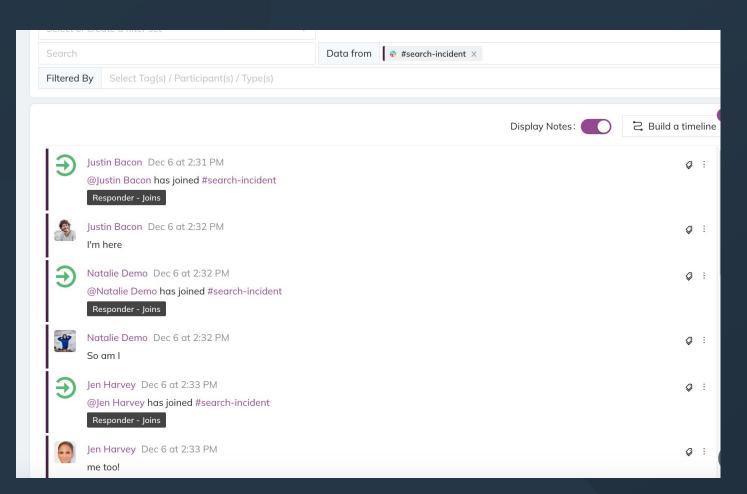












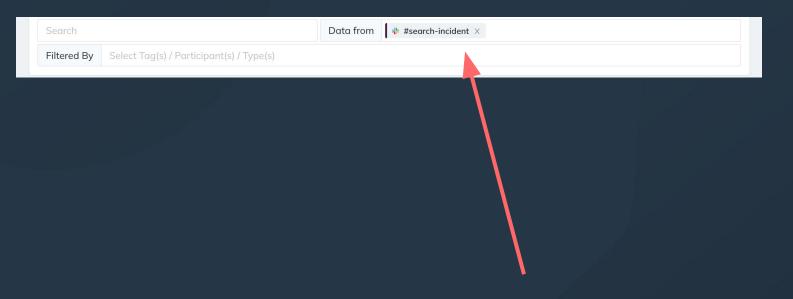




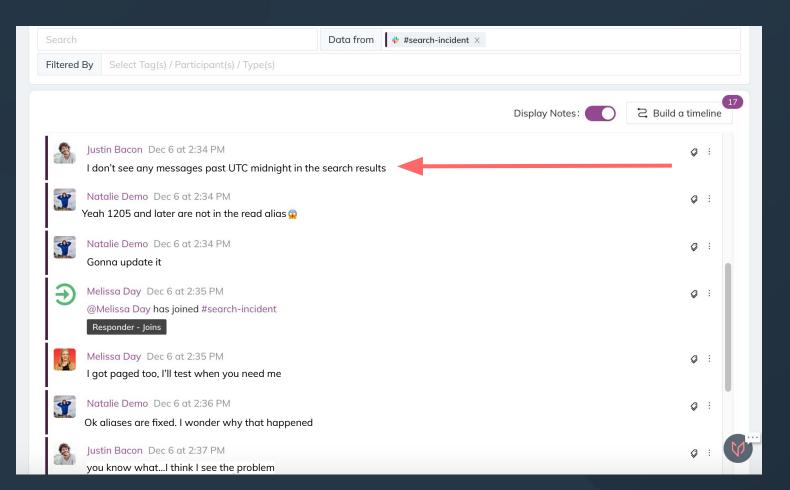
Conor Jacobs Customer Support |

Tenure	over 6 years Jul 11, 2016
Location	new york, ny Eastern Standard Time (UTC-05:00)
On call?	Permanently
Schedules	•Default

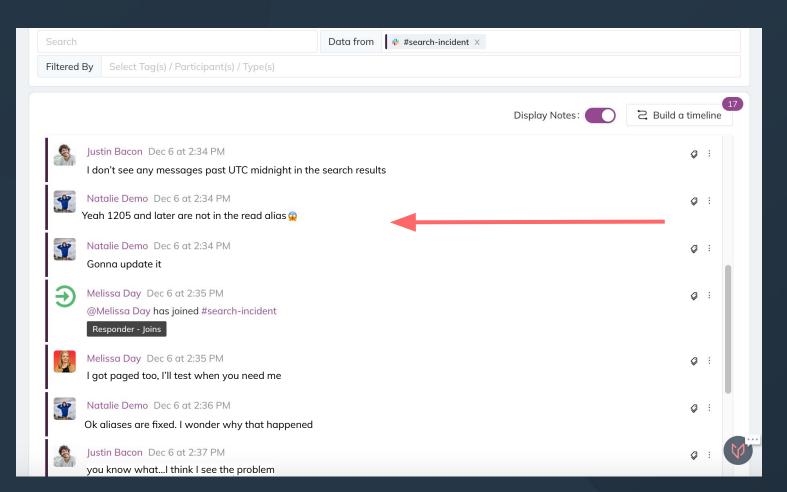




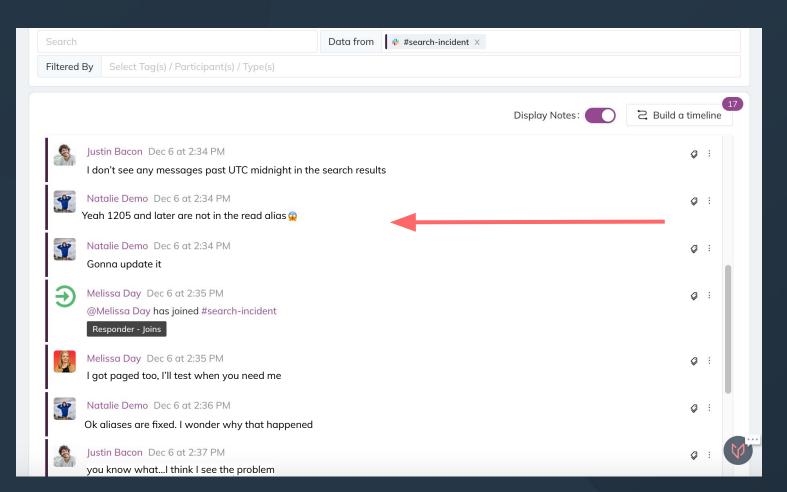




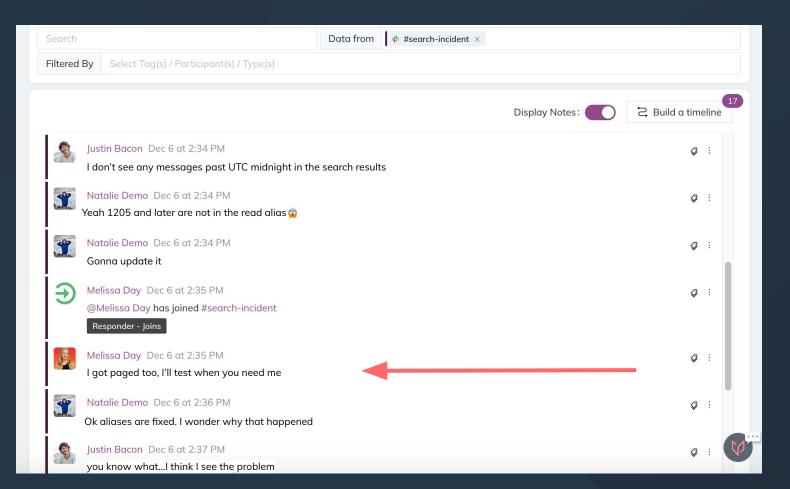




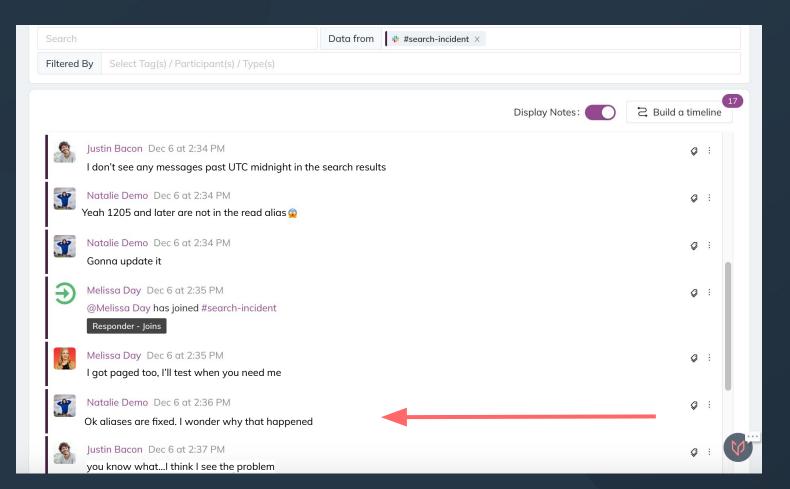




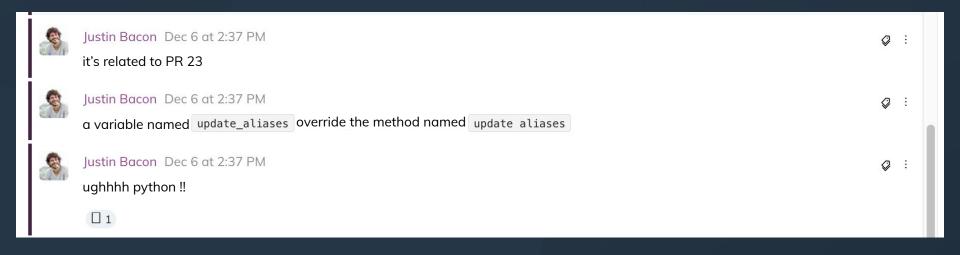








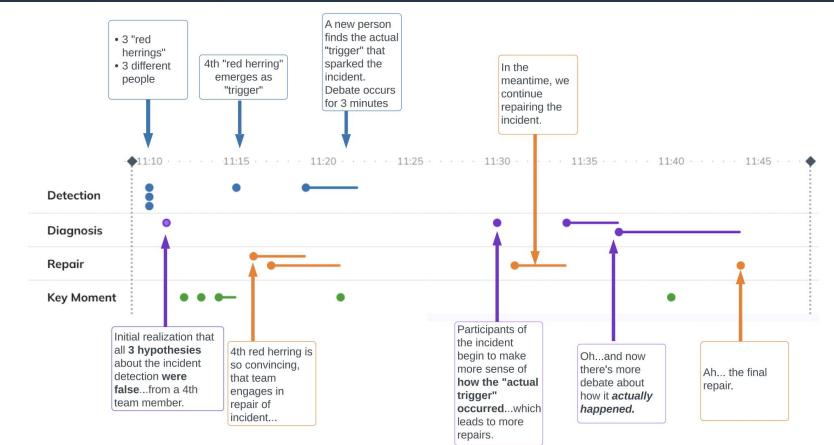














Incident #2: Key Takeaways

- There are multiple notions of what it means for search to be "working"
- Incident handling excellence is influenced by team cohesiveness and collaboration
- Changing code with alerts has risk
- Ease of coordination due to familiarity of team
- Quick awareness of issue at hand



Incident #2: Key Takeaways

- There are always multiple sharp-ends to study, and multiple people to invite to the meeting.
- An incident review can be used to gain insight into other's sharp ends.



	Incident 1	Incident 2	
Time to investigate	45 minutes	1 week	
Time of incident	20 minutes	Dates back to 3 weeks before the event	
Contributing Factors	2	8	
Action Items that got completed	2	6	
People involved	2	8	
Difficulties During Handling	No section	Dedicated section	
Impact	Customer	Customer + Employee	
Readers of the document	10	140	Mjeli

Part 3 - What's next



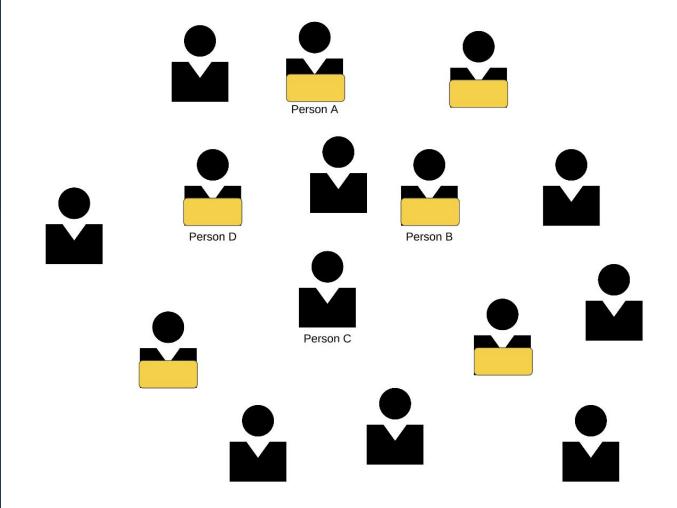
Working to understand how experts think, creates more experts.



Working to understand how experts think, creates more experts and helps them understand each other.

This is how you scale these people.







Tips for learning more from incidents:

- Use more than "customer impact" to warrant a review.
- Make it engaging create a record.
- Schedule and make time for learning. Don't break your promises to learn.
- Put the responsibility of running an incident review outside of the expert. Leadership needs to support this.

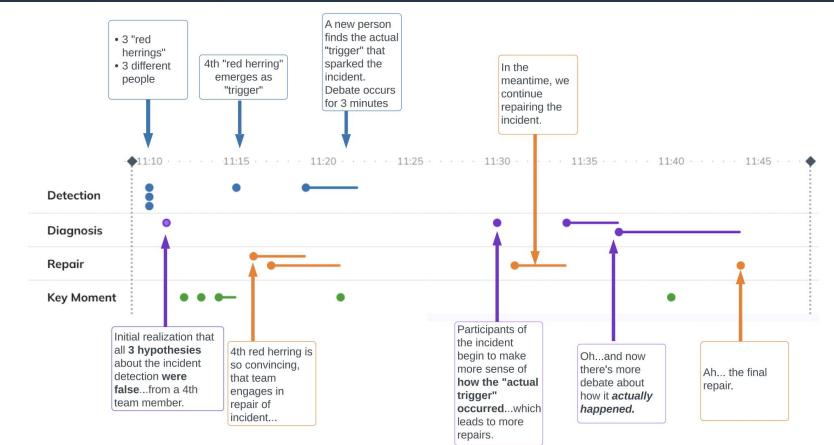


Tips for learning more from incidents:

- Use more than "customer impact" to warrant a review.
- Make it engaging create a record.
- Schedule and make time for learning. Don't break your promises to learn.
- Put the responsibility of running an incident review outside of the expert. Leadership needs to support this.



Make it engaging





Doing this can...

- Give the writer of the document more expertise
- Become training documents,
- Inform chaos experiments,
- Serve as professional development and refresher training,
- Help enable meta-analysis across incidents,
- Highlight how goal conflicts and changing organizational priorities,
- Identify where additional headcount is or is not needed,
- Inform build vs buy decisions,
- Create more experts.



Your metrics are wrong and they are lying to you. Stop making decisions based on them.



When you become an expert on learning from incidents, don't do it alone. Bring others with you.



Challenge for you: Attend or run a post-incident review in the next 2 weeks. I guarantee you can find something to analyze.



Keep the conversation going. Come find me!

Thank you

To learn more on how to unearth expertise:

https://www.jeli.io/howie nora@jeli.io

